



Development and clinical implementation of tailored image analysis tools for COVID-19 in the midst of the pandemic: The synergetic effect of an open, clinically embedded software development platform and machine learning

Constantin Anastasopoulos^{a,*}, Thomas Weikert^a, Shan Yang^b, Ahmed Abdulkadir^{c,d,e}, Lena Schmülling^a, Claudia Bühler^a, Fabiano Paciolla^a, Raphael Sexauer^a, Joshy Cyriac^b, Ivan Nesic^b, Raphael Twerenbold^f, Jens Bremerich^a, Bram Stieltjes^{a,b}, Alexander W. Sauter^a, Gregor Sommer^a

^a Department of Radiology, University Hospital Basel, University of Basel, Basel, Switzerland

^b Department of Research and Analysis, University Hospital Basel, University of Basel, Basel, Switzerland

^c Department of Old Age Psychiatry and Psychotherapy, Universitäre Psychiatrische Dienste Bern (UPD), University of Bern, Bern, Switzerland

^d Center for Biomedical Image Computing and Analytics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

^e Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

^f COVID-19 Research Coordinator, Department of Cardiology, University Hospital Basel, University of Basel, Basel, Switzerland

ARTICLE INFO

Keywords:

Computed tomography
COVID-19
Machine learning
Software

ABSTRACT

Purpose: During the emerging COVID-19 pandemic, radiology departments faced a substantial increase in chest CT admissions coupled with the novel demand for quantification of pulmonary opacities. This article describes how our clinic implemented an automated software solution for this purpose into an established software platform in 10 days. The underlying hypothesis was that modern academic centers in radiology are capable of developing and implementing such tools by their own efforts and fast enough to meet the rapidly increasing clinical needs in the wake of a pandemic.

Method: Deep convolutional neural network algorithms for lung segmentation and opacity quantification on chest CTs were trained using semi-automatically and manually created ground-truth ($N_{\text{total}} = 172$). The performance of the in-house method was compared to an externally developed algorithm on a separate test subset ($N = 66$).

Results: The final algorithm was available at day 10 and achieved human-like performance (Dice coefficient = 0.97). For opacity quantification, a slight underestimation was seen both for the in-house (1.8 %) and for the external algorithm (0.9 %). In contrast to the external reference, the underestimation for the in-house algorithm showed no dependency on total opacity load, making it more suitable for follow-up.

Conclusions: The combination of machine learning and a clinically embedded software development platform enabled time-efficient development, instant deployment, and rapid adoption in clinical routine. The algorithm for fully automated lung segmentation and opacity quantification that we developed in the midst of the COVID-19 pandemic was ready for clinical use within just 10 days and achieved human-level performance even in complex cases.

Abbreviations: CT, computed tomography; COVID-19, Coronavirus disease 2019; DCNN, deep convolutional neural network; FTE, full-time equivalent; HU, Hounsfield unit; PCR, polymerase chain reaction; POL, percentual opacity load; AI, artificial intelligence; A1-A3, algorithms 1-3.

* Corresponding author at: Department of Radiology, University Hospital Basel, University of Basel, Petersgraben 4, 4031, Basel, Switzerland.

E-mail addresses: constantinos.anastasopoulos@usb.ch (C. Anastasopoulos), thomas.weikert@usb.ch (T. Weikert), shan.yang@usb.ch (S. Yang), ahmed.abdulkadir@upd.unibe.ch (A. Abdulkadir), lenahannachristine.schmuelling@usb.ch (L. Schmülling), claudia.buehler@usb.ch (C. Bühler), fabiano.paciolla@usb.ch (F. Paciolla), raphael.sexauer@usb.ch (R. Sexauer), joshy.cyriac@usb.ch (J. Cyriac), ivan.nesic@usb.ch (I. Nesic), raphael.twerenbold@usb.ch (R. Twerenbold), jens.bremerich@usb.ch (J. Bremerich), bram.stieltjes@usb.ch (B. Stieltjes), alexander.sauter@usb.ch (A.W. Sauter), gregor.sommer@usb.ch (G. Sommer).

<https://doi.org/10.1016/j.ejrad.2020.109233>

Received 19 May 2020; Received in revised form 8 July 2020; Accepted 17 August 2020

Available online 28 August 2020

0720-048X/© 2020 The Author(s).

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Despite knowledge of the spread of the disease in Asia, Europe was overwhelmed by the dynamic of the new coronavirus disease (COVID-19) outbreak in spring 2020. Initial attempts to prevent its spread by geographically confined lockdowns failed and western countries became alerted by the exponential increase of new infections in Italy, which by the end of March had exceeded those reported by China. In mid-March, the European countries gradually entered a systemic lockdown and urgently prepared their healthcare systems for the challenges to come.

Driven by initial reports from China that indicated a higher sensitivity of chest computed tomography (CT) compared to polymerase chain reaction (PCR) in epidemic areas [1,2] imaging was recognized as an important additional diagnostic tool in the wake of the pandemic [3,4]. As a consequence, not only emergency and intensive care units but also radiology departments in Europe had to quickly adapt to the new reality, following recommendations from their colleagues in Asia [5].

In our department, a substantial increase in chest CT admissions for COVID-19 was seen soon after the initial cases were diagnosed in our hospital by the end of February. At first, CT was used for differential diagnosis of flu-like symptoms that had been advocated by early reports from China [6,7]. Soon, however, when the number of patients on the dedicated medical wards increased, our department received inquiries for a standardized method allowing quantification and follow-up of disease burden for supporting both triage towards intensive care and therapy decisions.

At the time of the initiation of this study, only little evidence was available on the evolution of lung tissue alterations in the course of the disease [8,9] and methodological proposals for quantification of these changes were at a very early stage [10,11]. Meanwhile the steadily growing literature on this topic is complemented by several other publications, including recent reports on visual scoring systems [12,13], first quantitative and deep convolutional neural network (DCNN) approaches [14–16] and a multicenter initiative for automated diagnosis and quantitative analysis of COVID-19 on imaging has been set up (<https://imagingcovid19ai.eu>).

This article describes the process from a prototypical development of automated artificial intelligence (AI) based software for lung segmentation and quantification of lung opacities in CTs of COVID-19 patients in a Research and Development environment to its clinical implementation within ten days. We discuss major strengths and weaknesses of our approach and set our results into the context of the current literature.

The underlying hypothesis was that modern academic centers in radiology are capable of developing and implementing a clinically useful AI based software for quantification of pulmonary opacities in COVID-19 by their own efforts and with sufficient speed to meet the rapidly increasing clinical needs in the wake of a pandemic.

2. Material and methods

2.1. Hospital and involved personnel

Our hospital is an academic hospital supplying maximum-care to a metropolitan area of 600.000 inhabitants. For this study, the Department of Radiology was supported by the Department of Research and Analysis, a team of 15 researchers with skills in image processing, data pipelines, deep learning applications and statistics. In the framework of this project, 3 members of latter department joined the newly created study team of 8 physicians, including 6 residents and 2 staff physicians specialized in cardio-thoracic imaging. The total work force dedicated to the project during the time frame of 10 days added up to approximately 2 full-time equivalent (FTE) for the scientists and 4 FTE for the physicians.

2.2. Patients and datasets

The prospective collection and evaluation of data from subjects with COVID-19 for this project was approved by the local ethics committee (approval number 2020-00566) as part of a study registered on ClinicalTrials.gov on 04/29/2020 (Identifier: NCT04366765). Data from patients actively denying consent for further research use were excluded. Finally, 152 datasets of COVID-19 patients with positive PCR were included (belonging to 146 patients), 23 performed with and 129 without iodine contrast administration, respectively. CT scans were performed in our institution on six scanners of four different types (Somatom Force, Edge, Definition Flash, and Definition AS+, all Siemens, Forchheim, Germany). Iterative image reconstruction (ADMIRE/SAPHIRE, Siemens Healthineers, Erlangen, Germany) with soft tissue kernel (I26f) was used.

Fig. 1 shows the increase in the cumulative number of patients from the beginning of March until the end of April 2020 ($N = 272$). For the training of the two algorithms A1 and A2 as defined below, 45 and 86 datasets were included, respectively, whereby the individual numbers depended on the availability of pre-processed datasets at the time of the training:

- Subset 1 consisting of 45 semi-automatically generated lung segmentations from the early phase of the project (until 03/27/2020) built the basis for training of a preliminary DCNN (A1).
- The second Subset of 86 manually edited segmentations of COVID-19 patients, including manual segmentations on Subset 1 scans (collectively named Subset 2_C, acquired until 03/30/2020) and accounting for half of the training subset for the final DCNN (A2).

Final testing was performed on a subset of 66 manually-edited segmentations on scans of COVID-19 patients acquired between 03/31/

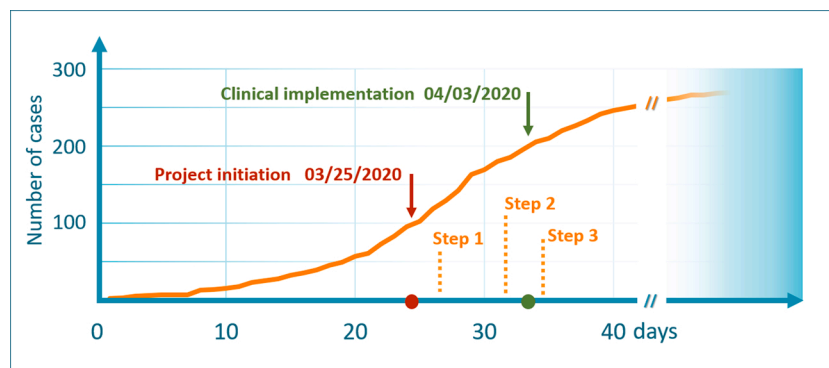


Fig. 1. Milestones of our project and cumulative sum of chest CT scans performed in patients with COVID-19 at our department plotted against time in the early pandemic period (day 1 to day 50).

2020 and 04/14/2020 (Subset 2_T).

In addition to the data of COVID-19 patients, 141 CT datasets of patients with different medical history other than COVID-19 were included. These were:

- Eighty-six chest CT datasets (including 16 follow-up examinations) that had been performed for assessment or exclusion of pulmonary infection during the second half of 2019 (Subset 2_{NC}). These were used together with the 86 scans of Subset 2_C in the training of algorithm A2.
- Fifty-five chest CT datasets from 2020 with different acquisition protocols (with and without intravenous contrast), used for validating the quantification of pulmonary opacities (see Appendix E₁ for further details).

The processing of this training and validation subsets was also approved by the local ethics committee (approval number 2020-00595).

Patient characteristics for the included datasets are given in Table S₁ of Appendix E₁.

2.3. Development pathway

The technical development, refinement and testing of methods followed a stepwise approach as listed below and visualized in Fig. 1.

2.3.1. Starting point – provisional, semi-automated assessment

First, evaluation with a semi-automated software package was introduced as an interim solution to provide ad hoc quantification of pulmonary opacities identified on CT ([10], CT Pulmo 3D, Syngo.via, VB30A, Siemens Healthineers, Erlangen, Germany). Originally designed for quantification of emphysema, this tool delineates lung contours with restricted options for manual correction. Ten equally spaced threshold-based subranges of negative Hounsfield units (HU) were formed. Based on previous reports on density distribution in acute respiratory distress syndrome [17], the percentage of voxels with HU values between -600 and 0 relative to the entire lung was calculated and considered during radiological reporting. The CT-scans from this initial evaluation step processed until 03/26/2020 (N = 45) are in the following referred to as Subset 1.

2.3.2. Step 1 – training of baseline in-house segmentation method

The 3D segmentations of Subset 1 were exported as binary masks without further manual intervention. For the training of deep learning algorithm 1 (A1), a framework for DCNN semantic segmentation with a U-Net architecture was used [18]. The data was processed with two convolutions in three spatial dimensions (3 × 3 × 3 convolution kernels). The principle of the network was based on the 3D U-Net without batch normalization [19] but we implemented it with only three resolution layers formed by two pooling and upsampling layers to reduce the model complexity and facilitate training. The number of channels per layer were the same as in the first three resolution levels in [19]. The training was performed on TensorFlow with NiftyNet (<https://niftynet.io>) on a consumer-grade graphics processor unit.

2.3.3. Step 2 – refinement of model by training with manual reference segmentations

The manually refined 3D segmentations from this step (n = 238) are in total referred to as Subset 2, which consists of the previously described Subsets 2_C (N = 86), 2_{NC} (N = 86), and 2_T (N = 66). Manually segmented reference standards of lung borders were generated in a medical image viewing/processing and software development platform *Nora* (www.nora-imaging.com). After discussing the segmentation strategy, three radiologists in training (T.W., L.S., C.B.) segmented the lung borders in each axial slice of the chest CT dataset using a threshold-based annotation pencil and avoiding inclusion of hilar vessels. From the Subset 2_T, 10 cases were randomly selected for inter-rater comparisons

of independent segmentations by all three raters, from which a single estimated ground truth segmentation was computed, as in [16].

The same network architecture as for A1 (Step 1) was used for training of algorithm A2, trained with Subset 2_C and Subset 2_{NC} (total N = 172).

2.3.4. Step 3 – implementation of a third-party lung segmentation algorithm

On 04/04/2020 an independent research group released “COVID-19 Web” on the *GitHub* platform, which to our knowledge was the first open-source lung segmentation algorithm specifically trained with COVID-19 chest CT datasets (<https://github.com/JoHof/lungmask>) [15]. Similar to A1 and A2, it is based on the U-Net architecture and had been trained on 40 and 238 datasets from patients with and without COVID-19, respectively (1–6 ratio). We implemented version 0.2.2 (downloaded on April 7th, in the following referred to as A3) in step 3 as an external reference for our own algorithms A1 and A2.

2.4. Data evaluation and statistical analysis

The inference of lung borders was followed by a simple post-processing step for all three algorithms, during which spurious remote segmentations were excluded, while keeping the largest connected components. Algorithms A1–3 were tested and compared on Subset 2_T.

2.4.1. Segmentation performance

Descriptive statistics were used to compare established performance metrics comparing inter-rater and human/deep learning whole-lung segmentations:

- The Dice similarity coefficient ranging from 0 to 1, as the number of common voxels times two, divided by the sum of voxels from each segmentation
- The maximal Hausdorff distance in mm, as the maximum distance between two segmentation contours

The Dice coefficient was additionally evaluated for the upper and lower 20 % of the lungs (Appendix E₁, Data Supplement S₃).

2.4.2. Threshold-based quantification

We estimated the percentual opacity load (POL) in both lungs by thresholding between -600 and 0 HU:

$$POL_{-600} = \sum_{s=1}^n \frac{O_s}{L_s} \times 100$$

where O voxel count of lung mask with $-600 \leq HU \leq 0$ and L voxel count of lung mask in slice s , respectively.

POL_{-600} derived from each of the algorithms were separately compared to manual POL_{-600} in Subset 2_T with Bland-Altman analyses in R (v 3.6.3) [20].

Additionally, quantification was computed in 55 non-COVID-19 cases in varying acquisition phases after administration of intravenous iodine contrast (Appendix E₁, Data Supplement S₄).

2.5. Implementation and availability

After the pathway development, a data pipeline was set up to navigate acquired images from the scanner to *Nora*, where the proposed algorithm (A2) was implemented. This locally hosted software is available to the radiologist during reporting through a web browser. Upon arrival of the image dataset, a segmentation and quantification process are operated automatically or can be triggered by mouse-click. In verified COVID-19 cases or in patients with a high pretest probability, the results are exported to the clinical PACS in the form of a visual report and included into the radiological report at the decision of the radiologist and are thus accessible to the attending physicians (see video).

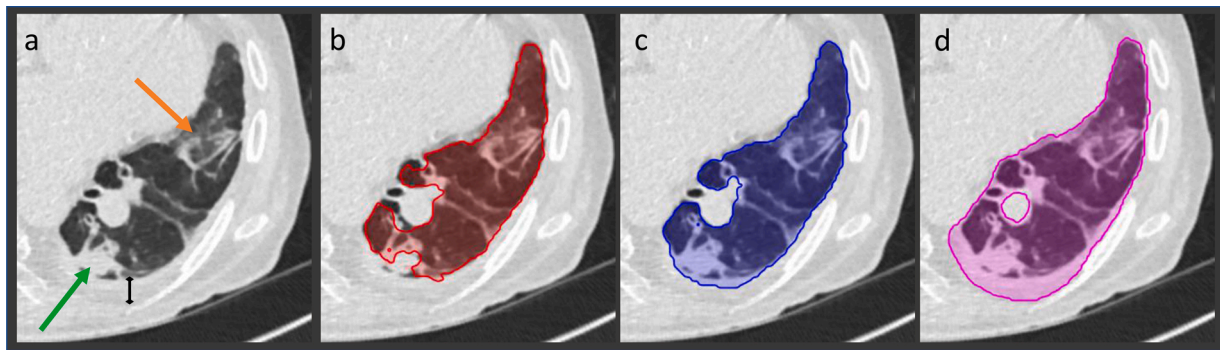


Fig. 2. Segmentation examples of algorithms A1-A3: left basal lung (transversal slice) in an atypical case of COVID-19 (a) with ground-glass opacities (orange arrow), consolidations (green arrow) and a pleural effusion (black line). (b): lung borders, including ground-glass opacity but not consolidation, are segmented with algorithm A1. (c): lung border segmentation including both the ground-glass opacity and the consolidation with algorithm A2. (d): pleural effusion is unexpectedly included in the lung border segmentation with the third-party algorithm A3.

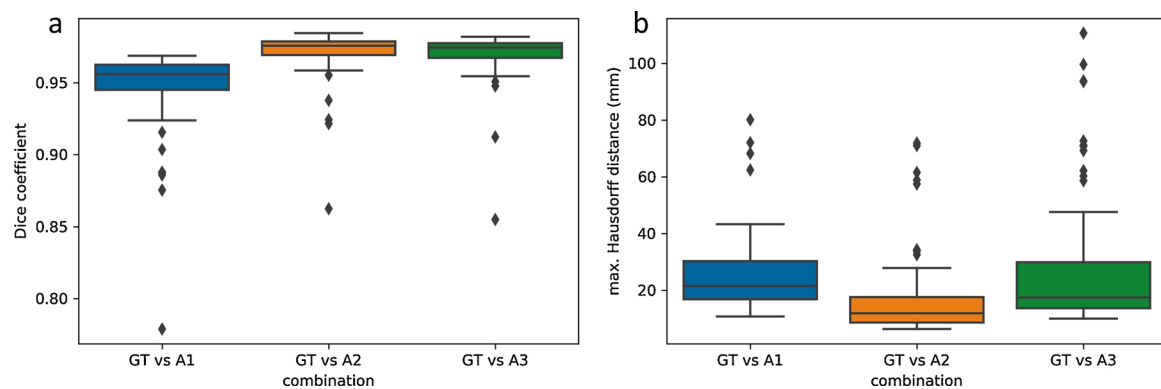


Fig. 3. Boxplots of Dice coefficient (a) and maximal Hausdorff distance (b) for the three algorithms (blue: algorithm A1, orange: algorithm A2 and green: algorithm A3), compared to the manual ground truth (GT) on the test subset. The lowest outlier in the Dice coefficient of all three algorithms occurred in one case with a one-sided pneumothorax.

3. Results

3.1. Development pathway

The technical development steps 1 through 3 were successfully completed within a period of 10 days: At day 1 and 2, segmentations generated by the semi-automated approach were exported and converted to a suitable format for the training of algorithm A1. The pre-processing, training of A1 and subsequent processing took place on day 3–4. Data preparation for the training of algorithm A2 including thorough manual segmentation was performed on days 5–8, followed by training of A2. The preliminary results of the deep learning segmentation A2 were available and discussed on day 9. As a side note, the total number of chest CT scans acquired for COVID-19 at our institution had in the meantime exceeded 200. On day 10 we finalized the threshold-based quantification and the complete pipeline was uploaded to our clinically embedded software development platform *Nora*. Fig. 2 shows a comparison of the output of the automated approach of algorithms A1-A3. The detailed timeline of the development pathway is given in Table S₂ of Appendix E₁.

3.2. Segmentation performance for whole-lung

The performance metrics are given in Fig. 3 and Table 1. The precision of the deep learning lung tissue segmentation in Subset 2_T was excellent for A2 and A3 with mean Dice coefficients of 0.97, while A1 showed a slightly lower mean Dice coefficient of 0.95. The maximal Hausdorff-distance showed a mean of 25, 17 and 28 mm for A1, A2 and A3, respectively. Isolated outliers were observed above the upper

quartile, mainly for the preliminary (A1) and the third-party (A3) algorithm, corresponding to the unexpected inclusion of pneumothorax or pleural effusion in the lung segmentation (for an example see Fig. 2). Inter-rater segmentation comparison on the 10 cases was excellent with mean Dice coefficients of 0.99 for all comparisons (Table 1).

3.3. Performance in opacity quantification

The results of the threshold-based quantification analysis are displayed in Fig. 4 and Table 2. Lung opacities were diagnosed in forty-four out of sixty-six scans (66 %) from the test dataset 2_T. POL_{-600} of the manually lung segmentations ranged from 5 to 55%. The algorithms showed mean underestimations of 3.0 %, 1.8 % and 0.9 % for A1, A2 and A3 in Subset 2_T, respectively. For A1 and A3, there was a proportional bias in POL_{-600} towards higher opacity loads (0.35 % and -0.4 % for each 10 % increase in POL_{-600} , respectively), while for A2 the slope of the bias was almost zero (Fig. 4).

3.4. Implementation and availability

Implementation has been accomplished during the pandemic in April 2020. In the first 6 weeks after implementation, almost 500 chest CTs admitted to our department were processed automatically. These scans were in part from patients with COVID-19 ($N \approx 100$) and in part from patients with other diseases from the later period of the COVID-19 pandemic. Processing time per individual is approximately 5 min. The quantification results stored in the PACS are used by the attending physicians as a risk stratification tool, particularly as an indicator of deterioration urging transfer to intensive care.

Table 1

Descriptive statistics for performance metrics Dice coefficient and maximum Hausdorff distance, on the left for comparisons between each algorithm and the human reference standards and on the right for the inter-rater comparisons.

GT vs algorithm comparison				Inter-rater comparison			
		Dice coefficient	max. Hausdorff distance		Dice coefficient	max. Hausdorff distance	
Mean	GT vs A1	0.95	25.5	r1 vs r2	0.99	17.0	
	GT vs A2	0.97	17.4	r1 vs r3	0.99	22.0	
	GT vs A3	0.97	28.4	r2 vs r3	0.99	23.7	
SD	GT vs A1	0.03	14.2	r1 vs r2	0.01	8.7	
	GT vs A2	0.02	15.2	r1 vs r3	0.01	20.8	
	GT vs A3	0.02	24.7	r2 vs r3	0.01	21.4	
Minimum	GT vs A1	0.78	10.8	r1 vs r2	0.97	7.5	
	GT vs A2	0.86	6.3	r1 vs r3	0.97	7.5	
	GT vs A3	0.86	10.0	r2 vs r3	0.97	4.9	
Median	GT vs A1	0.96	21.5	r1 vs r2	0.99	14.6	
	GT vs A2	0.98	11.8	r1 vs r3	0.99	15.6	
	GT vs A3	0.97	17.4	r2 vs r3	0.99	17.5	
maximum	GT vs A1	0.97	80.1	r1 vs r2	1	34.7	
	GT vs A2	0.98	71.9	r1 vs r3	1	78.3	
	GT vs A3	0.98	111.0	r1 vs r2	1	78.3	

Abbreviations: GT: ground truth, SD: standard deviation, A1-A3: algorithms 1–3, r1-r3: rater1–3.

4. Discussion

In this work, the steps required for the development of a deep learning algorithm for quantification of opacities on chest CT and its clinical integration are outlined. Within just 10 days, two deep neural network segmentation algorithms with different sizes of training datasets were trained and a threshold-based quantification approach estimating lung opacity load was implemented into the clinical workflow in the midst of the pandemic. The results were compared to manually processed results and a lung segmentation algorithm developed and released in the same time period by an independent research group.

The successful and timely implementation of the described pipeline was partially owed to the prior experience of our team in the implementation of deep learning pipelines and the close collaboration between developers and clinical stakeholders. In contrast to academic benchmark challenges in image processing, where usually a minimum of three months is provided for fine-tuning of the model, an end-to-end machine learning pipeline with incremental steps and sanity checks was applied, in fact a well-established practice in the tech industry [21]. Additionally, the seamless implementation of the third-party algorithm in the software development platform *Nora* allowed us a head-to-head comparison in real-time and is an additional strength of this study. The availability of open-source algorithms is vital for the field, especially in time-critical occasions, therefore our proposed method was made available on an open-source platform.

The first deep learning algorithm A1 was deployed as early as 4 days after project initiation. Forty-five chest CTs from patients with COVID-19 available at that time were utilized for its training. Even without prior manual refinement of the segmentations, A1 provided a satisfactory overall segmentation performance. Nevertheless, a significant progress from the preliminary results of A1 to the second algorithm A2 was seen. This latter deep learning algorithm, trained with the fourfold amount of chest CT datasets and after manual refinement of the reference standard, showed a considerable improvement in segmentation performance on the test subset, consisting of chest CTs with a wide spectrum of COVID-19 related opacities. Although comparable to the third-party algorithm (A3), A2 showed in average lower maximal Hausdorff distances, pointing at a better agreement with the reference standard segmentation. This might be in part attributed to the fact that the reference standard of training and test subsets were created by the same human raters [22], but is also reflected by higher segmentation accuracy in specific cases with coexisting pneumothorax or massive pleural effusion.

The deep learning segmentation was interlinked with a subsequent opacity quantification step, based on a HU-thresholding method,

established at our site for COVID-19 related opacities before the deep learning segmentations were introduced. Quantification based on thresholding has been previously used for differentiation of normal lung tissue from opacities, such as ground-glass or consolidations [17] and the lower cut-off of -600 HU implemented here reflects the counterpart of the “well-aerated” lung, which has been correlated to the severity of the disease and clinical outcome in patients with COVID-19 [14]. The average underestimation of automated quantification was minor for the proposed A2 and the third-party algorithm, whereas the latter showed a negative bias slope towards higher opacity loads. In contrary, quantification bias with A2 did not manifest a dependency on opacity load, thus making the estimation of error from automated quantification more predictable when comparing baseline and follow-up scans.

The approach proposed in this article quantifies but does not classify lung opacities, as recently shown by an automated differentiation of lung opacities in chest CTs caused by COVID-19 and acquired pneumonia [16]. Direct segmentation of affected lung areas has also been proposed as an alternative to approaches using thresholding after segmentation, although the voxel misclassifications reported there might eventually result in a similar degree of opacity underestimation [23]. Factors influencing the distribution of HU values in lung tissue such as inflation depth and prior contrast administration were identified in a small sample with no lung opacities (Appendix E₁ only). The role of these and other factors, such as partial volume effects from vessels and bronchi, common CT artifacts or coexisting lung disease, have to be evaluated in more detail.

The presented pragmatic approach also harbors some limitations. The first is the selection and curation of datasets that was strongly dominated by the question of availability of pre-processed data from the provisional, semi-automated pipeline in the wake of the pandemic. The data that was used for this project does therefore neither represent a complete, consecutively acquired sample of the COVID-19 cohort at our hospital nor is it a fully random sample. In addition, the reference standard subset of COVID-19 chest CT for the training of A2 was extended by an equally sized subset of relatively homogeneous scans performed for exclusion of pulmonary infection with and without iv contrast. Taken together, these inconsistencies in data selection may limit the performance of the algorithm in cases of advanced COVID-19, although for the training subset of the third-party A3 algorithm this portion of chest CTs performed for other reasons than COVID-19 was even lower. On the other hand, this augmentation of the training subset might reduce the selection bias and represents a more diverse sample [24]. An additional limitation is the use of empirical HU-thresholds for disease quantification, since clear-cut thresholding, especially the upper threshold of 0, might fail to include dense lung consolidations or might

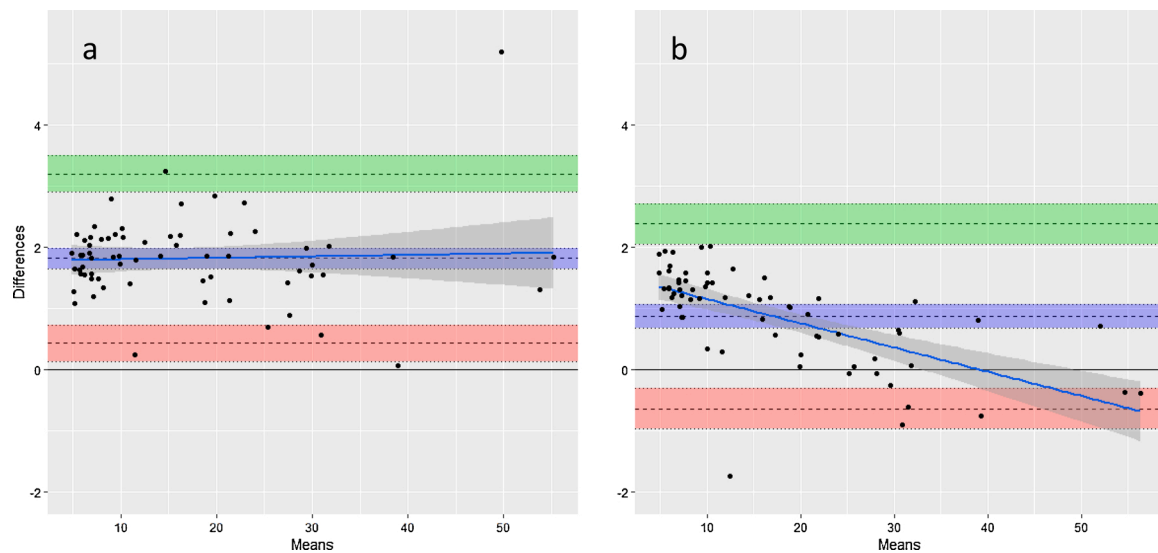


Fig. 4. Bland-Altman analyses for opacity quantification (POL₆₀₀, in %) derived from manual and deep learning segmentations (a: manual vs A2, b: manual vs A3). POL₆₀₀ values in the test subset of patients with COVID-19 ranged from 5 to 55 %. Note that numbers below 7% do not necessarily reflect lung opacities but can also be found in healthy lungs (see Data Supplement S4).

Table 2

Bland-Altman analyses of opacity quantification (in %) between the manual reference standard and each of the 3 algorithms.

Algorithm	Mean bias	Lower limit of agreement	Upper limit of agreement	Proportional bias intercept	Proportional bias slope
A1	3.1	1.2	5.0	2.5	0.035
A2	1.8	0.4	3.2	1.8	0.002
A3	0.9	-0.7	2.4	1.5	-0.04

A1 to A3: deep neural network algorithms1–3.

erroneously include small vessels due to partial volume effects. Finally, future contributions to the deep learning algorithm have to take into account coexisting abnormalities and potentially include more datasets from a wider spectrum of diseases in the training and test subsets in a controlled fashion.

5. Conclusion

Rapid development of reliable lung segmentation for COVID-19 is feasible with DCNN. Clinically acceptable results can be achieved in only few days and with less than 50 cases available for training, while by fourfold increasing of the number of datasets human-level performance was achieved. With this technique, fully automated quantification of pulmonary involvement in COVID-19 is possible even in the presence of advanced disease with extensive consolidations. This demonstrates the potential of including machine learning to assist clinical processes to manage the current pandemic and beyond.

CRedit authorship contribution statement

Constantin Anastasopoulos: Conceptualization, Methodology, Software, Visualization, Writing - original draft. **Thomas Weikert:** Methodology, Data curation, Writing - original draft. **Shan Yang:** Software, Investigation. **Ahmed Abdulkadir:** Software, Investigation. **Lena Schmülling:** Data curation, Validation. **Claudia Bühler:** Data curation. **Fabiano Paciolla:** Data curation. **Raphael Sexauer:** Data curation. **Josh Cyriac:** Software. **Ivan Nesic:** Software. **Raphael Twerenbold:** Validation. **Jens Bremerich:** Supervision. **Bram Stieltjes:** Supervision, Writing - review & editing. **Alexander W. Sauter:** Conceptualization, Supervision, Writing - review & editing. **Gregor Sommer:**

Conceptualization, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgement

The article processing charge was funded by the University of Basel in the funding programme Open Access Publishing.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ejrad.2020.109233>.

The algorithm and reporting tool can be downloaded from following link: doi:<https://doi.org/10.5281/zenodo.4012205>.

References

- [1] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of chest CT for COVID-19: comparison to RT-PCR, *Radiology* (2020), 200432, <https://doi.org/10.1148/radiol.20200432>.
- [2] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, et al., Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, *Radiology* (2020), 200642, <https://doi.org/10.1148/radiol.20200642>.
- [3] M.-P. Revel, A.P. Parkar, H. Prosch, M. Silva, N. Sverzellati, F. Gleeson, A. Brady, COVID-19 patients and the radiology department - advice from the European Society of Radiology (ESR) and the European Society of Thoracic Imaging (ESTI), *Eur. Radiol.* (2020), <https://doi.org/10.1007/s00330-020-06865-y>.
- [4] G.D. Rubin, C.J. Ryerson, L.B. Haramati, N. Sverzellati, J.P. Kanne, S. Raoof, et al., The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the fleischner society, *Radiology* (2020), 201365, <https://doi.org/10.1148/radiol.202001365>.
- [5] J. Wu, J. Pan, D. Teng, X. Xu, J. Feng, Y.-C. Chen, Interpretation of CT signs of 2019 novel coronavirus (COVID-19) pneumonia, *Eur. Radiol.* (2020), <https://doi.org/10.1007/s00330-020-06915-5>.
- [6] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, et al., CT imaging features of 2019 novel coronavirus (2019-nCoV), *Radiology* 295 (2020) 202–207, <https://doi.org/10.1148/radiol.20200230>.
- [7] H.X. Bai, B. Hsieh, Z. Xiong, K. Halsey, J.W. Choi, T.M.L. Tran, et al., Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT, *Radiology* (2020), 200823, <https://doi.org/10.1148/radiol.20200823>.
- [8] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z.A. Fayad, N. Zhang, et al., Chest CT findings in coronavirus Disease-19 (COVID-19): relationship to duration of infection, *Radiology* (2020), 200463, <https://doi.org/10.1148/radiol.20200463>.
- [9] Y. Wang, C. Dong, Y. Hu, C. Li, Q. Ren, X. Zhang, et al., Temporal changes of CT findings in 90 patients with COVID-19 pneumonia: a longitudinal study, *Radiology* (2020), 200843, <https://doi.org/10.1148/radiol.20200843>.

- [10] L. Tang, X. Zhang, Y. Wang, X. Zeng, Severe COVID-19 pneumonia: assessing inflammation burden with volume-rendered chest CT, *Radiology: Cardiothoracic Imaging* 2 (2020), e200044, <https://doi.org/10.1148/ryct.2020200044>.
- [11] Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, H. Shi, Longitudinal assessment of COVID-19 using a deep learning-based quantitative CT pipeline: illustration of two cases, *Radiology: Cardiothoracic Imaging* 2 (2020), e200082, <https://doi.org/10.1148/ryct.2020200082>.
- [12] M. Prokop, W. van Everdingen, T. van Rees Vellinga, J. van Quarles Ufford, L. Stöger, L. Beenen, et al., CO-RADS - A categorical CT assessment scheme for patients with suspected COVID-19: definition and evaluation, *Radiology* (2020), 201473, <https://doi.org/10.1148/radiol.2020201473>.
- [13] R. Yang, X. Li, H. Liu, Y. Zhen, X. Zhang, Q. Xiong, et al., Chest CT severity score: an imaging tool for assessing severe COVID-19, *Radiology: Cardiothoracic Imaging* 2 (2020), e200047, <https://doi.org/10.1148/ryct.2020200047>.
- [14] D. Colombi, F.C. Bodini, M. Petrini, G. Maffi, N. Morelli, G. Milanese, et al., Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia, *Radiology* (2020), 201433, <https://doi.org/10.1148/radiol.2020201433>.
- [15] Hofmanninger J., Prayer F., Pan J., Rohrich S., Prosch H., Langs G. Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem, 31-Jan-20, 10 pp. <http://arxiv.org/pdf/2001.11767v1>.
- [16] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, *Radiology* (2020), 200905, <https://doi.org/10.1148/radiol.2020200905>.
- [17] L. Gattinoni, P. Caironi, P. Pelosi, L.R. Goodman, What has computed tomography taught us about the acute respiratory distress syndrome? *Am. J. Respir. Crit. Care Med.* 164 (2001) 1701–1711, <https://doi.org/10.1164/ajrccm.164.9.2103121>.
- [18] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, *Proceedings of MICCAI* (2015) 234–241.
- [19] Ö Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: *Proceedings of MICCAI*, 9901, 2016, pp. 424–432.
- [20] Datta Deepankar, Deepankardatta/Blandr: Version 0.5.1 - Cran Submission Release: Zenodo, 2017.
- [21] N. Polyzotis, S. Roy, S.E. Whang, M. Zinkevich, Data lifecycle challenges in production machine learning, *SIGMOD Rec.* 47 (2018) 17–28, <https://doi.org/10.1145/3299887.3299891>.
- [22] J. Mongan, L. Moy, C.E. Kahn, Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers, *Radiology: Artificial Intelligence* 2 (2020), e200029, <https://doi.org/10.1148/ryai.2020200029>.
- [23] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, L. Xia, Serial quantitative chest CT assessment of COVID-19: deep-learning approach, *Radiology: Cardiothoracic Imaging* 2 (2020), e200075, <https://doi.org/10.1148/ryct.2020200075>.
- [24] M.P. Recht, M. Dewey, K. Dreyer, C. Langlotz, W. Niessen, B. Prainsack, J.J. Smith, Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations, *Eur. Radiol.* (2020), <https://doi.org/10.1007/s00330-020-06672-5>.